Received Date : 07-Mar-2016
Revised Date : 24-May-2016
Accepted Date : 14-Jun-2016
Article type : Original Article

Contrasting support for alternative models of genomic variation based on microhabitat preference: species-specific effects of climate change in alpine sedges

Authors: Rob Massatti[1,2] and L. Lacey Knowles[1]

[1]Department of Ecology and Evolutionary Biology, The University of Michigan, Ann Arbor MI, USA 41809-1079

[2]Corresponding Author: Rob Massatti, E-mail: massatti@umich.edu

Key words: *Carex*, Cyperaceae, glaciation, Pleistocene, Rocky Mountains

Contact Information (same address for both):
Lacey Knowles, knowlesl@umich.edu, 734-763-5603
Rob Massatti, massatti@umich.edu, 307-920-1028
University of Michigan
2085 Museums Building
1109 Geddes Ave.
Ann Arbor, MI 48109-1079

Word count: 6985
Number of Figures: 4
Number of Tables: 1
Data archive: Dryad Digital Repository - doi:10.5061/dryad.ng3bv

ABSTRACT

Deterministic processes may uniquely affect co-distributed species' phylogeographic patterns such that discordant genetic variation among taxa is predicted. Yet, explicitly testing expectations of genomic discordance in a statistical framework remains challenging. Here, we construct spatially and temporally dynamic models to investigate the hypothesized effect of microhabitat preferences on the permeability of glaciated regions to gene flow in two closely related montane species. Utilizing environmental niche models from the Last Glacial Maximum and the present to inform demographic models of changes in habitat suitability over time, we evaluate the relative probabilities of two alternative models using approximate Bayesian computation (ABC) in which glaciated regions are either (i) permeable or (ii) a barrier to gene flow. Results based on the fit of the empirical data to datasets simulated using a spatially explicit coalescent under alternative models indicate that genomic data are consistent with predictions about the hypothesized role of microhabitat in generating discordant patterns of genetic variation among the taxa. Specifically, a model in which glaciated areas acted as a barrier was much more probable based on patterns of genomic variation in *Carex nova*, a wet-adapted species. However, in the dry-adapted *C. chalciolepis*, the permeable model was more probable, although the difference in the support of the models was small. This work highlights how statistical inferences can be used to distinguish deterministic processes that are expected to result in discordant genomic patterns among species, including species-specific responses to climate change.

INTRODUCTION

Understanding the contribution of species-specific attributes to observed patterns of genetic variation is critical for determining why taxa responded similarly (or dissimilarly) to historical climate changes.. Fortunately, newly developed, spatially explicit methodologies that generate genetic expectations under alternative scenarios (e.g., Neuenschwander *et al.* 2008; Knowles & Alvarado-Serrano 2010; Brown & Knowles 2012) offer an opportunity to evaluate the relative support for competing hypotheses related to species' life history characteristics. For example, comparing the fit of empirical data to models that capture differences in species' habitat requirements or in the habitat stability across a landscape (e.g., He *et al.* 2013) may provides insight into the role of biotic factors in structuring genetic variation. However, given that the number of biologically informed models that could be tested for any given study system is limitless, the challenge is how to decide what models to test (Knowles 2009).

Even though it might be possible to compare a hundred different models (e.g., Pelletier & Carstens 2014), such an approach may not be desirable. For example, if the differences among the models are trivial (e.g., the models differ in what may be considered nuisance parameters that do not impact their interpretation), the biological insights provided by selecting one model over another will be limited. Such inherent constraints of model-based approaches reinforce the importance of developing models that illuminate processes of biological interest (Papadopoulou & Knowles 2015a, 2016; but see O'Meara *et al.* 2015 for a dissenting point of view). In other words, it is not only the analytical approach per se, but the creativity and intimate knowledge of a study system that a researcher brings to such tests that ultimately determines how much insight a model-based phylogeographic analysis might provide. Rather than building narratives centered upon how well empirical data should fit generic scenarios (e.g., testing for a correlation under an isolation by distance model), insights can be gained by developing narratives derived from organisms' natural histories, as well as historical ecological and climatic factors, to generate testable hypotheses (Papadopoulou & Knowles 2016). Given that not all biological characteristics (e.g., dispersal capability or habitat specificities) of a species may be important, or that their effects may vary depending upon the geographic or temporal scale of study (Papadopoulou & Knowles 2016), the key is to identify those characteristics that provide insights about the processes structuring genetic variation (e.g., the role of geographic barriers or changing

climatic conditions). When coupled with data simulated under biologically informed models, the relative fit of empirical genetic data to alternative models forms the basis for rigorous statistical tests of competing hypotheses (see also Bruggeman *et al.* 2010; Epperson *et al.* 2010; Landguth *et al.* 2010; Morgan *et al.* 2011; Shirk *et al.* 2012; Papadopoulou & Knowles 2015b).

Here we test the role of species' microhabitat preferences and their potential interactions with shifting distributions associated with climatic changes in structuring the geographic distribution of genetic variation in two species of montane sedges (*Carex* L., Cyperaceae). Briefly, although the species co-occur within montane habitat across the southern Rocky Mountains, *C. nova*, an inhabitant of wet microhabitats (i.e., occurring predominantly within drainages) may have been disproportionately displaced to lower elevations (relative to dry-adapted species) because of the large accumulation of snow and ice in this microhabitat during glacial periods (Fig. 1). As a consequence of population persistence being limited to lower elevations around the margins of glaciated habitat, populations of *C. nova* may have been relatively isolated during the glacial periods that predominated throughout the Pleistocene (i.e., most of the time during the Pleistocene is represented by glacial, not interglacial, periods). This narrative contrasts with plant species that inhabit meadows, ridges, and slopes, such as *C. chalciolepis*. Inhabitants of drier microhabitats may have persisted in high elevation areas that remained free of glaciers and persistent snow throughout glacial cycles (although the persistence of such populations is controversial; Wachter *et al.* 2016), in addition to populations that may have established at lower elevations. As a result, *C. chalciolepis* populations may have remained more interconnected throughout glacial periods. Although the geographic patterning of population structure observed in these two species is consistent with such a narrative (as described in Massatti & Knowles 2014), evaluating the extent to which differences in the patterns of genetic variation can be ascribed to microhabitat differences requires formalization of these narratives into models for statistical testing.

To test whether current genetic structure reflects a species' ability/inability to persist within higher elevation microhabitats during glacial periods, we generated genetic expectations under two alternative models – one with glaciated regions as a permeable habitat and another with the regions as a barrier to gene flow – using the integrative distributional, demographic, and coalescent (iDDC) modeling approach (He *et al.* 2013). Specifically, we created demographic models that incorporated spatial and temporal heterogeneity in climatically suitable areas, as

informed by present and past ecological niche models, and required that populations either persist within or be excluded from glaciated areas (see also Currat & Excoffier 2004; Wegmann *et al.* 2006). By performing tests of the fit of the empirical data to simulated datasets using approximate Bayesian computation (ABC; see Beaumont *et al.* 2002 for an overview of ABC) and tests of model validation, we evaluated whether the species differ in their support for the alternative models in a manner consistent with the hypothesis that microhabitat differences determine how species respond to shifts in climate. This study is not only a detailed analysis of how species-specific properties may determine whether taxa respond similarly (or dissimilarly) to climate change, but our work also illustrates how general narratives about the processes structuring genetic variation can be formalized into models for statistical testing. We discuss our findings in terms of our ability to generalize the effects of climate change on montane communities as well as acknowledge general caveats with our analyses that leave open questions requiring further analysis.

MATERIALS AND METHODS

Genomic data generation and processing

Carex chalciolepis (N = 110) and *C. nova* (N = 109) leaf material was field collected from 11 sampling localities across the full extent of their geographical ranges in the southern Rocky Mountains (Fig. 1; Table S1). Within populations, the collecting distance among individuals was maximized to decrease the probability of sampling related individuals (average distance between samples of 300 m, and a minimum distance of 35 m). Leaf material was stored in silica gel until DNA was extracted with DNeasy Plant Mini Kits (Qiagen, Hilden, Germay) following the manufacturer's protocol. As with previous libraries (see Massatti & Knowles 2014), anonymous genomic loci were developed using a restriction associated DNA sequencing (RADseq) approach (for details see Peterson *et al.* 2012); library construction and data processing is described in full detail in the Supporting Information. Briefly, fragments ranging in size from 400 to 500 base pairs were sequenced at The Centre for Applied Genomics (Hospital for Sick Children, Toronto, Canada) to generate 50 base pair, single-end reads. Single nucleotide polymorphisms (SNPs) were identified using a multinomial-based likelihood model that

accounts for sequencing error implemented in Stacks v1.25 (Hohenlohe *et al*. 2010; Catchen *et al*. 2011; Catchen *et al*. 2013).

Five Illumina 2500 sequencing runs were used to generate data for this project. In order to i) maximize the number of unlinked loci, ii) reduce missing data to the fullest extent, iii) maximize the number of individuals per population, and iv) ensure that the subsampled SNP datasets displayed the same genetic patterns among populations as the larger, unfiltered datasets, we employed the following post-processing procedures. Only RADseq loci containing up to three SNPs were retained, and for each RADseq locus, one randomly selected SNP was exported into a STRUCTURE-formatted file if the locus contained less than 50% missing data. Patterns of genome-wide SNP variation among individuals and populations were then visualized for each species with PCAs in the program R (R Core Team 2014) using the 'adegenet' package (Jombart 2008) and the 'dudi.pca' function; missing data were replaced by the mean frequency of the corresponding allele. Subsequently, we minimized missing data by manually removing SNPs and individuals containing an excess of missing data and rechecking PCAs to ensure that the subsampling procedure did not alter the major axes of genetic variation among populations (see also Huang & Knowles 2014). Individuals of *C. chalciolepis* and *C. nova* contained an average of 4.9% and 5.3% missing data, respectively (Table S2).

Custom scripts were used to convert the STRUCTURE-formatted files into ARLEQUIN-formatted files, which were input into ARLSUMSTAT to extract the empirical summary statistics used in ABC (see below). We also used the empirical ARLEQUIN-formatted files to create masks that were applied to the simulated datasets so the amount and pattern of missing data in the simulated data would match precisely that of the empirical datasets. All custom scripts and genomic data are deposited in Dryad (doi:10.5061/dryad.ng3bv).

iDDC approach

Simulated genetic datasets to evaluate the support for the alternative models – that is, (i) glaciers as barriers, and ii) permeable glaciated regions (which allowed for local persistence within and gene flow across glaciated regions) (Fig. 2) – were generated using the iDDC approach (described in detail in He *et al.* 2013). Briefly, simulated genetic datasets were generated using a coalescent model informed by a spatially and temporally explicit demographic model that reflected differences in habitat suitabilities across space and time (based on the

ENMs; described below), in addition to either population persistence within or exclusion from glaciated areas (i.e., the hypothesized predictions for the dry and wet-adapted species, respectively). The fit of the empirical genetic data to the two alternative models was evaluated using approximate Bayesian computation (ABC), along with procedures to validate parameter estimates and model quality (Wegmann *et al.* 2010). All scripts for the iDDC analyses are deposited in Dryad (doi:10.5061/dryad.ng3bv).

Quantification of habitat suitability. Habitat suitability across the southern Rocky Mountains during the present and LGM was estimated for *C. chalciolepis* and *C. nova* with MAXENT v3.3.3e (Phillips *et al*. 2006). Nineteen bioclimatically informative variables for the present (WorldClim v1.4; Hijmans *et al*. 2005) and the LGM (PMIP2-CCSM; Braconnot *et al.* 2007) were used to generate the environmental niche models (ENMs; full details are presented in the Supporting Information). Georeferenced distribution points representative of each species' entire range were collected from personal fieldwork and validated voucher specimens housed at the Rocky Mountain Herbarium (species distribution points are available at doi:10.5061/dryad.ng3bv). To have a computationally tractable number of cells for demographic simulations (detailed below), we statistically downscaled the cell sizes of the ENMs to 0.42 decimal degrees (~16.5 km^2 per cell) (e.g., Ray *et al*. 2010; He *et al.* 2013). Subsequently, the values of the cells in the LGM and present ENMs denoting the logistic habitat suitability scores (ranging continuously from 0 to 1) determined by MAXENT were reassigned. Specifically, the logistic values were grouped into ten categories using the 'equal interval' clustering method in ArcMap 10.0 (ESRI, Redlands, California, USA) and assigned values ranging from 1-10. Given the extreme similarity of the ENMs between the species (for both the present and past; Fig. S1), an average of the habitat suitability scores of the two species was generated for each time period and used for the demographic simulations (see details below). By using estimates of habitat suitabilities for the present and past based on averages for the two species, we provide a standardized model that avoids the confounding influences of subtle differences in the unique ENMs of the taxa that could contribute to the relative probabilities of the alternative models (as opposed to differences in the permeability of the glaciated areas themselves, which is the central focus of the study). In addition to the present and LGM ENMs that were generated for each species, landscapes of habitat suitabilities specific to the two alternative models were generated: one in which the

glaciated areas were a barrier and one in which these regions remained permeable (detailed below), corresponding to the hypothesized effect of differences in microhabitat (i.e., glaciated regions would have been permeable in the dry, but not the wet-adapted species). Note that because microhabitat differences are manifest at a small spatial scale (i.e., meters or less) due to the complex interaction between topography and environmental conditions in montane regions, the "habitat" suitabilities of the taxa based on quantification from the ENMs are similar and only convey that environmental conditions are suitable for the species occurrence.

Simulated datasets. Demographic and coalescent simulations were performed in SPLATCHE2 (Ray *et al.* 2010) to generate patterns of genetic variation where habitat suitabilities differed across a landscape and through time (see Knowles & Alvarado-Serrano 2010; Brown and Knowles 2012). For each of the two models tested (i.e., glaciers as barriers vs. permeable habitat; Fig. 2), 1 000 000 simulated datasets were generated for each species separately (i.e., 4 000 000 total). Uniform priors were used for the demographic parameters (i.e., between-deme migration rate, *m*, maximum carrying capacity of a deme, *K*, and the population sizes of the initial populations, N_{Anc}), and the carrying capacities of demes were scaled proportionally to their habitat suitability values. In other words, data were simulated over a range of maximum *K*-values, with the highest quality habitat (demes with values of 10) reaching full carrying capacity (100%), while carrying capacities of demes associated with lower quality habitat were decreased proportionally (see Knowles & Alvarado-Serrano 2010; Fig. 2). Each generation, *m* proportion of the population migrates out of the local deme to adjacent cells (i.e., to the north, south, west, and east). After the exchange of individuals, local demes grow logistically at the rate of 1, regulated by the carrying capacity inferred from the habitat suitability. Demographic simulations were initialized with a southern, central, and northern population (see Fig. 2), each with a population size of N_{Anc}. Note that population trees estimated using empirical SNP data for *C. chalciolepis* and *C. nova* clustered populations sampled from these regions, validating our use of three initial populations.

The landscapes that informed the demographic modeling procedure differed over time. Specifically, for both the barrier and permeable models, an initial landscape based on the ENM at the LGM was used to inform the demographic simulations from generations 1-750. The initial landscape was followed by a model-specific landscape (i.e., a landscape representing either (i) glaciated areas as barriers or (ii) permeable glaciated areas) from generations 751-1750, which

corresponded to the height of the last Pleistocene glaciation. For the final period from generations 1751-2083, the demographic modeling was informed by a landscape with habitat suitabilities based on an ENM estimated using contemporary climatic conditions for both models (Fig. 2). With a generation time of 3 years for these high elevation species (Körner 2003) and the scaling of the generations by a factor of 15 (to make simulations computationally tractable), the simulations span a time period from the present to the LGM and beyond. Note that modeling was extended beyond the LGM (i.e., the first 750 generations) to provide a sufficient amount of time for all suitable habitat to be occupied over all possible combinations of population demographic parameters. Because of this scaling, any biological interpretation of absolute values of population genetic parameters would need to be adjusted accordingly.

As noted above, the initial and final landscapes used to perform the demographic simulations were the same between the models (and species), whereas the landscape corresponding to the height of the LGM for the two models differed in one key aspect: the habitat suitability values of the glaciated areas (which were identified from maps of glacial moraines and glacial till; see Ehlers & Gibbard 2004; Colorado Geological Survey http://coloradogeologicalsurvey.org/). Specifically, under the model in which glaciers would have acted as barriers, demes reconstructed as glaciated were assigned a *K*-value of 0. In contrast, in the alternative model in which such glaciated areas were permeable, the *K*-values were decreased by 85% of the original value estimated from the LGM ENM, which corresponds to impeded dispersal relative to the surrounding non-glaciated areas (Fig. 2), with a lower bound of $K = 20$. This lower bound was used because of the uncertainty surrounding habitat quality estimates in demes where habitat suitability scores fell below the maximum training sensitivity plus specificity threshold identified by MAXENT. Varying the *K*-values for demes with highly uncertain habitat suitability could introduce demographic consequences that would have undue influence on the resulting patterns of genetic variation. Preliminary analyses confirmed that gene flow was able to occur among populations separated by glaciers despite reduced carrying capacities. Because *C. chalciolepis* and *C. nova* were predicted to occur in and adjacent to montane habitat during the LGM, but not in the geographically distant lower elevation basins and plains (which were predominantly represented by demes with values of 1), all demes with habitat suitability values of 1 were assumed to be uninhabitable (see Fig 2).

A spatially explicit coalescent model informed by the deme-specific demographic parameters was used to simulate genetic data (i.e., genetic variation differed across the landscape depending on the specific combination of *m*, *K*, and N_{Anc}; Excoffier *et al*. 2000; Currat *et al.* 2004). An independent coalescent process was run to generate a genealogy for each locus analyzed in the empirical data (i.e., 1142 and 1010 coalescent simulations for *C. chalciolepis* and *C. nova*, respectively) for each of the 4 000 000 datasets simulated with different combinations of demographic parameters (the range of which as specified by the priors). Each simulated dataset consisted of the same number of individuals from the same the corresponding sampling localities, with the same amount of missing data, as in the empirical data.

For each of the simulated genetic datasets, as with the empirical data, nine summary statistics were calculated using ARLSUMSTAT v.3.5.2 (Excoffier & Lischer 2010). These included the number of segregating sites (*S*) for each population and across populations, mean heterozygosity across loci for each population and across populations (*H*), and pairwise population F_{ST} (Weir & Cockerham 1984), for a total of 83 summary statistic values calculated per simulated dataset.

Model selection and validation. Approximate Bayesian computation (ABC) was used to select between alternative models, as implemented with ABCestimator in ABCtoolbox (Wegmann *et al*. 2010). Rather than using the 83 summary statistic values independently to estimate parameters, partial least squares (PLS) components (Boulesteix & Strimmer 2007) were extracted from the summary statistics using the "PLS" package (Mevik & Wehrens 2007) with boxcox treatment (Box & Cox 1964) in R for the first 20 000 simulations for each model and species. This approach removes the effects of interactions between summary statistics and reduces "the curse of dimensionality" associated with using a large number of summary statistics (Boulesteix & Strimmer 2007). We examined the root mean squared error (RMSE) prediction for each parameter to decide how many PLS components to use for parameter estimation (Fig. S2). Of the million simulated datasets generated for each species under a particular model, 5000 simulations (0.5%) whose summary statistics were closest to those calculated from the empirical genomic data were retained and used for parameter estimation and model selection. Post-sampling regression adjustment was applied using the ABC-GLM (general linear model) function (Leuenberger & Wegmann 2010) in R to obtain posterior distributions of the

parameters. Bayes factors, which are the ratios between marginal densities of the two models, were used for model selection; a higher ratio indicates more support for the first model (Jeffreys 1961).

To evaluate whether a model is capable of generating data similar to the empirical data, the likelihood of the empirical data was compared to the likelihoods of the retained simulations under the GLM model. If all the retained simulations have a better likelihood than the observed data (i.e., a low *P*-value), it would indicate a model is highly unlikely (Wegmann *et al*. 2010). A coefficient of variation (R^2) of each parameter explained by the PLS components was also computed and used as an indicator of the power of estimation (Neuenschwander *et al.* 2008). For the most probable model selected for each species, the accuracy of parameter estimates was validated using 1000 pseudo-observations generated from prior distributions of the parameters. If estimation of the parameters is unbiased, posterior quantiles of the parameters from pseudo runs should be uniformly distributed (Cook *et al*. 2006; Wegmann *et al.* 2010). The posterior quantiles of true parameters for each pseudo run were also calculated based on the posterior distribution of the regression-adjusted 5000 simulations closest to the pseudo-observation.

RESULTS

Empirical genomic dataset

Almost 180 000 000 (average 1 630 280 ±719 977 per individual) and 172 000 000 (average 1 579 380 ±733 329 per individual) reads were generated for the 110 and 109 *C. chalciolepis* and *C. nova* individuals, respectively (Table S2). The retention of reads after data processing and assembly with Stacks averaged 86% per individual for both species. The final datasets contained 1142 loci with 1 SNP per locus across 101 individuals of *C. chalciolepis* and 1010 loci with 1 SNP per locus across 99 individuals of *C. nova*, after post-processing steps to remove loci and/or individuals because of missing data. The datasets had good representation of all sampled populations for tests with the iDDC approach (Table S3).

Model selection and validation

Based on the relative information content contained in the PLS components, the first six PLSs of the predictor variables (i.e., the summary statistics) were retained for ABC analyses (see

Fig. S2). Posterior distributions of parameter estimates were distinct from the prior, indicating they contained information relevant to estimating the parameters (Fig. 3). Moreover, comparison of the posterior distributions before and after GLM regression adjustment of the 5000 retained simulations shows a marked improvement of parameter estimates with the GLM regression (Fig. 3). The accuracy of the parameter estimates varied (Table 1). Specifically, in addition to a flatter posterior probability (Fig. 3), the lowest power for estimating parameter values was associated with the ancestral population size (N_{Anc}), as indicated by the lowest R^2 values across models (Table 1). Nevertheless, tests of potential bias of parameter estimates show that posterior distributions of N_{Anc} are uniformly distributed in both species (Fig. 4). In contrast, the histograms of the posterior quantiles of *m* and *K* did deviate significantly from a uniform distribution for both species based on analyses of 1000 pseudo-observed datasets, suggesting a potential bias (Fig. 4).

Based on the marginal densities calculated from the 5000 retained simulations for each model, the model with the best fit to the empirical data (based on Bayes factors) differed between species. Specifically, the model with glaciated areas as barriers was more probable for *C. nova*, whereas the model with permeable glaciated areas was more probable for *C. chalciolepis* (although in the latter case, the difference was marginal), which corresponds to the hypothesized effect of microhabitat differences (Table 1). Evaluation of whether the most probable model was capable of producing the empirical data in each species (i.e., consideration of the *P*-values) indicated that not only were they a good fit, but they were a better fit (i.e., a larger *P*-value) than the alternative models, suggesting a better correspondence between the empirical data and the simulated data (Table 1).

DISCUSSION

Our study highlights the formalization of biologically informed hypotheses into a statistical model-testing framework (see Papadopoulou & Knowles 2016) to evaluate whether species-specific traits may mediate the effects of climate change. Specifically, we show that species' responses to glaciations may be intricately tied to their microhabitat preferences. Below we discuss the implications of our findings when considering the effects of climate change on co-distributed taxa, and especially on species in montane communities.

In contrast to studies that rely upon correlative approaches to identify factors that may structure genetic variation (see Massatti & Knowles 2014), here the link between genetic patterns and process is explicit (i.e., data were simulated with either population persistence within or exclusion from glaciated areas – the hypothesized predictions for the dry and wet-adapted species, respectively, based on the differential accumulation of glaciers on ridges and drainages; see Fig. 1) and the fit of empirical data to alternative models was statistically evaluated. Moreover, our study differs from other model-based approaches in which a generic or "naïve" model is applied (i.e., a model that is not informed by biological knowledge of a particular system) (Pelletier & Carstens 2014; Grummer *et al.* 2015; O'Meara *et al*. 2015); we designed our study to target species that differed in one key trait – microhabitat. Minimizing differences between taxa allowed us to explicitly test hypotheses aimed at providing specific insights about the contribution of species-specific traits to patterns of genetic variation (albeit with some caveats; see below). In addition to discussing the utility of the iDDC approach to test if phylogeographic discord reflects deterministic factors, we address both the challenges and limitations with such inferences. Lastly, we highlight the insights that comparative phylogeography can provide regarding the role of biotic factors in structuring genetic variation.

Ecological and evolutionary implications for montane taxa

The finding of a correspondence between a model in which glaciated areas were barriers to gene flow and genomic variation in the wet-adapted *C. nova*, as opposed to a model in which glaciated areas were permeable, as with genomic variation in the dry-adapted *C. chalciolepis* (Table 1), has a number of ecological and evolutionary implications. From an ecological perspective, our work highlights how generalizations about the response of montane communities to climate change may be overly simplified (see also Alexander *et al.* 2016). Importantly, the structure of genomic variation of the *Carex* species differs significantly (i.e., the parameterization and fit of the alternative models differed between the taxa; Table 1), despite both sedges being common and co-distributed (Massatti *et al.* 2016) and having many biological similarities, including traits associated with dispersal capabilities. Moreover, the respective fit of the dry-adapted versus wet-adapted species to a model with either population persistence within or exclusion from glaciated areas is in line with expectations derived from knowledge of the interactions of persistent snow and ice with microhabitats in montane ecosystems (Ehlers &

Gibbard 2004). As such, this is the first study to directly test causal mechanisms related to the persistence of populations in glaciated regions (Lohse *et al.* 2011; Westergaard *et al.* 2011; Lanier *et al.* 2015; Wachter *et al.* 2016). Specifically, differing support for the most likely model between the species (Table 1) suggests that species adapted to wetter microhabitats were more isolated around the margins of glaciers, whereas species adapted to drier microhabitats persisted within glaciated regions (as well as establishing at lower elevations) and remained relatively connected by gene flow.

It is generally accepted that statistical evaluation of alternative processes that might have generated observed patterns of genetic variation (e.g., through model selection procedures, as used here) is critical because similar genetic patterns may result from different demographic processes (Csillery *et al.* 2010; He *et al.* 2013). However, our study is more than an example of model-based statistical phylogeography (Knowles 2009). Specifically, in comparison with other comparative phylogeographic analyses, our study stands out because a deterministic process that is expected to generate discordant genomic variation among taxa is evaluated (i.e., the fit of alternative models is predicted to differ depending on a taxon's microhabitat, corresponding to the hypothesized persistence in or exclusion from previously glaciated areas). This contrasts with the tradition of relying upon phylogeographic concordance for assessing the role of competing processes, and where discord is commonly attributed to the idiosyncrasies of history (reviewed in Papadopoulou & Knowles 2016). Nevertheless, there are some caveats with our approach. Specifically, while we have conducted analyses aimed at addressing the quality of our inference (discussed in the following section), there are additional hypotheses that we cannot rule out. Note that all model-based comparative phylogeographic studies face this issue (Knowles 2009) – that is, this caution is not unique to our study, or specific to the iDDC procedure per se. Moreover, from our perspective, such a discussion is extremely helpful for considering analyses that may be worth pursuing in the future to provide additional insights into the role of biotic factors structuring genetic variation. In this regard, we would argue that the merit of our study, and of comparative phylogeographic studies more generally, is not in identifying "the evolutionary history" of taxa. Instead, it is the insights gained by identifying the probability of one hypothesis relative to others, which is a function of the both the study design and its execution. As noted above, and given that patterns of genetic variation in many species are postulated to reflect the interactions of ecology and climatic oscillations (Avise & Walker 1998; Hewitt 2000; Carstens

& Knowles 2007), our results set the stage for additional work that is needed before any broad generalizations can be derived from model-based comparative phylogeographic analyses about the role of species-specific traits in structuring genetic variation (see also Papadopoulou & Knowles 2016).

Validation and interpretation of model-based inferences

There are both methodological and conceptual aspects of model-based inferences that must be considered to avoid erroneous conclusions or tests that offer limited insights. For example, there are a number of methodological issues that can be especially challenging with ABC procedures (see Oaks *et al.* 2013) and the iDDC approach we applied. These include issues with approximating the likelihood of models with summary statistics (Pritchard *et al*. 1999; Beaumont *et al.* 2002) as opposed to using all of the data as in full likelihood-based models (Hey & Nielsen 2004, 2007; Kuhner 2006; Nielsen & Beaumont 2009; Hey 2010). Moreover, in addition to difficulties associated with particular steps in ABC procedures (e.g., post-sampling adjustment when the relationship between parameters and summary statistics is extrapolated beyond the region of the observed data set; see Beaumont *et al*. 2002; Leuenberger & Wegmann 2010), ABC will always produce a posterior distribution, even if the model is a poor fit to the data (Bertorelle *et al*. 2010). Accordingly, model validation is critical.

Several approaches we applied suggest that our results are generally robust. We evaluated the potential bias in parameter estimates (Fig. 4), contrasted the posterior probability of parameter estimates pre and post-GLM (Fig. 3) and relative to the prior, as well as utilized the RMSE of parameter estimates (Fig. S2) to inform decisions about the inclusion of PLS components. While the predictive power of the data for some parameters differed (see R^2 in Table 1) and the posterior quantiles calculated from pseudo observed datasets of both *m* and *K* showed a significant departure from a uniform distribution (Cook *et al*. 2006; Wegmann *et al.* 2010), we note that for both species, the most probable model provided not only (i) a good fit to the empirical genomic data, but (ii) it was a better fit compared to the less probable model (Table 1). Specifically, the Bayes factor indicates strong support for the barrier model in *C. nova*, and this model has a much higher probability of generating simulations with likelihood-values comparable to the empirical data compared to the permeable model (Table 1). In *C. chalciolepis*, the marginal densities of the two models are much more similar. As a consequence, although the

empirical data is more probable under the permeable model, the difference in support based on the Bayes factor is not strong. Nevertheless, it is worth noting that, even though some combinations of parameters produce datasets that match the *C. chalciolepis* empirical data under the barrier model, the permeable model has a much wider parameter region that generates data close to the empirical data (as reflected in differences in the *P*-values for the two models; Table 1). The fit of the empirical data under these complex models is very encouraging because it can be difficult to capture the complicated nature of a species' history. For example, despite approaches for evaluating complex models, the likelihood of the empirical data under the most probable model may be much lower than any data simulated under such a model (see Excoffier *et al.* 2013).

Even though our models are capable of generating the data (Table 1), this does not mean the most probable models for the two species are the "correct" ones. We acknowledge that there could be other models not considered here that might fit the empirical data. However, this does not discount the insights gained with respect to the goal of the study, which was to test whether the empirical data of two species would support alternative models as predicted if the microhabitats of the species mediated their responses to climate change. In this regard, comparison of the estimated parameter values may illuminate possible differences in the population dynamics of the species under climate change scenarios. In particular, we note that *C. chalciolepis* tended toward higher values of K_{max} and lower values of *m* compared to *C. nova*, while the difference in N_{Anc} was more ambiguous because of uncertainty in the estimation of this parameter (Fig. 3). Specifically, the PLSs of the summary statistics were informative for all parameters in both species, but estimates of N_{Anc} are associated with only moderate R^2 values, in contrast to the very high R^2 for the other parameters (Table 1). In the context of glaciations, these combinations of parameter values may intimate that habitat stability for dry-adapted species facilitated larger effective population sizes and lower rates of migration, while continual disturbance within wet microhabitats fostered relatively lower population sizes and higher rates of migration in wet-adapted species (e.g., the reestablishment of populations in disturbed habitats). Our analyses support this proposition, but the detected biases in some parameter estimations (Fig. 4) cautions against interpreting the parameters directly (see Wegmann *et al.* 2010).

Evaluating potential demographic differences between the species under changing climatic conditions is an important area for future consideration. Such investigations may include tests of whether support for alternative models reflects differences in how estimates of habitat suitability (as informed by ENMs) scale to population demographic parameters. For example, a particular value of habitat suitability may not translate into equivalent predicted carrying capacities between *Carex* species (as modeled here). While there are not pronounced differences in the contemporary abundances of *C. chalciolepis* and *C. nova* that suggest a nonequivalent relationship between the suitability of a habitat and local population sizes, differences in the local stability of the species' populations, or even local adaptation, could potentially contribute to the different patterns of genetic variation in ways not explicitly accounted for in our models. Likewise, we do not directly model microhabitat preference per se, but instead test models with either population persistence or exclusion from glaciated areas based on the hypothesized predictions for the dry and wet-adapted species, respectively, based on the differential accumulation of glaciers on ridges and drainages (see Fig. 1). It is possible that factors other than microhabitat preference might contribute to the persistence or exclusion of taxa from glaciated areas. Hence, it may not be microhabitat preference, but possibly some untested co-varying explanatory variable, that drives the differences in the fit of the taxa to the alternative models. However, the similarity of the taxa makes it difficult to identify other hypothetical, yet realistic, factors. Moreover, any such hypothetical factor would not only have to result in contrasting support for the alternative models between the taxa, but also preserve the directionality of the model fits (e.g., *C. chalciolepis*, but not *C. nova*, must fit the model with persistence within glaciated areas, and vice a versa for the model with exclusion from glaciated areas). We note that other species that are closely related to and co-distributed with *C. chalciolepis* and *C. nova* (Massatti *et al.* 2016) will facilitate the exploration of these intriguing hypotheses, but such tests are beyond the scope of our present analyses.

Conclusions

Using intimate knowledge of the interactions between climate and topography within montane ecosystems, as well as utilizing data from other disciplines (e.g., maps of glacial till and glacial moraines; Fig. 1), we tested alternative models to elucidate the potential impact of glaciers on co-distributed species. We evaluated the relative fit of empirical data under a model

in which glaciated areas were a barrier versus one in which they were permeable to test the biologically informed hypotheses that differences in microhabitat preferences would result in predictable differences in the responses of the taxa to climate change. Our results supported the hypothesized predictions for the dry and wet-adapted species based on the differential accumulation of glaciers on ridges and drainages (Fig. 1) – the barrier model was the most probable for *C. nova*, whereas the model with permeable glaciated regions was more probable in *C. chalciolepis* (although in the latter case the difference was not strong) (Table 1).

The models and approach we apply here go beyond traditional analyses common in phylogeography (e.g., tests of isolation by distance; Slatkin 1993) and comparative phylogeography (e.g., relying on concordant patterns for inferring the role of factors in structuring genetic variation; Papadopoulou & Knowles 2015a). Such approaches are in their infancy and have only been applied in a limited number of studies (e.g., Neuenschwander *et al*. 2008; He *et al.* 2013; Martinkova *et al*. 2013). By combining the power that genomic data provide with the proper validation of complex models, approaches such as iDDC provide an exciting opportunity to address ecological and evolutionary principles in a comparative phylogeographic framework that cannot be addressed using traditional methodologies (Excoffier *et al.* 2013; Papadopoulou & Knowles 2016).

Acknowledgements

The authors are grateful for the support and helpful advice of their colleagues, especially Qixin He on analyses and Anna Papadopoulou for comments on an earlier draft of the manuscript. This work was funded by NSF (DEB 1309072 to RM) and the Hinsdale Scholarship from the University of Michigan Museum of Zoology to RM.

REFERENCES

Alexander JM, Diez JM, Levine JM (2016) Novel competitors shape species' responses to climate change. *Nature*, **525**, 515–518.

Avise JC, Walker D, Johns GC (1998) Speciation durations and Pleistocene effects on vertebrate phylogeography. *Proceedings of the Royal Society B: Biological Sciences,* **265**, 1707–1712.

Beaumont MA, Zhang WY, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025–2035.

Bertorelle G, Benazzo A, Mona S (2010) ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Molecular Ecology*, **19**, 2609–2625.

Boulesteix A-L, Strimmer K (2007) Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. *Briefings in Bioinformatics*, **8**, 32–44.

Box GEP, Cox DR (1964) An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **26**, 211–252.

Braconnot P, Otto-Bliesner B, Harrison S *et al.* (2007) Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum–Part 1: experiments and large-scale features. *Climate of the Past*, **3**, 261–277.

Brown JL, Knowles LL (2012) Spatially explicit models of dynamic histories: examination of the genetic consequences of Pleistocene glaciation and recent climate change on the American Pika. *Molecular Ecology*, **21**, 3757–3775.

Bruggeman DJ, Wiegand T, Fernandez N (2010) The relative effects of habitat loss and fragmentation on population genetic variation in the red-cockaded woodpecker (*Picoides borealis*). *Molecular Ecology*, **19**, 3679–3691.

Carstens BC, Knowles LL (2007) Estimating Species Phylogeny from Gene-Tree Probabilities Despite Incomplete Lineage Sorting: An Example from *Melanoplus* Grasshoppers. *Systematic Biology*, **56**, 400–411.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) *Stacks*: Building and Genotyping Loci *De Novo* From Short-Read Sequences. *G3 Genes, Genomes, Genetics*, **1**, 171–182.

Catchen J, Hohenlohe P, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology*, **22**, 3124–3140.

Cook SR, Gelman A, Rubin DB (2006) Validation of software for Bayesian models using posterior quantiles. *Journal of Computational and Graphical Statistics*, **15**, 675–692.

Csillery K, Blum MGB, Gaggiotti OE, Francois O (2010) Approximate Bayesian computation (ABC) in practice. *Trends in Ecology & Evolution*, **25**, 410–418.

Currat M, Ray N, Excoffier L (2004) SPLATCHE: a program to simulate genetic diversity taking into account environmental heterogeneity. *Molecular Ecology Notes*, **4**, 139–142.

Currat M, Excoffier L (2004) Modern humans did not admix with Neanderthals during their range expansion into Europe. *PLoS Biology*, **2**, 2264–2274.

Ehlers J, Gibbard PL eds. (2004) *Quaternary Glaciations - Extent and Chronology II: North America*. Elsevier, London, United Kingdom.

Epperson BK, McRae BH, Scribner K *et al.* (2010) Utility of computer simulations in landscape genetics. *Molecular Ecology*, **19**, 3549–3564.

Excoffier L, Dupanloup I, Huerta-Sanchez E *et al.* (2013) Robust Demographic Inference from Genomic and SNP Data. *PLoS Genetics*, **9**, e1003905.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Excoffier L, Novembre J, Schneider S (2000) SIMCOAL: a general coalescent program for the simulation of molecular data in interconnected populations with arbitrary demography. *Journal of Heredity*, **91**, 506–509.

Grummer JA, Calderón-Espinosa ML, Nieto-Montes de Oca A *et al.* (2015) Estimating the temporal and spatial extent of gene flow among sympatric lizard populations (genus *Sceloporus*) in the southern Mexican highlands. *Molecular Ecology*, **24**, 1523–1542.

He Q, Edwards DL, Knowles LL (2013) Integrative testing of how environments from the past to the present shape genetic structure across landscapes. *Evolution*, **67**, 3386–3402.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Hey J (2010) Isolation with migration models for more than two populations. *Molecular Biology and Evolution*, **27**, 905–920.

Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*, **167**, 747–760.

Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences, USA*, **104**, 2785–2790.

628 Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated
629 climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.
630
631 Hohenlohe PA, Bassham S, Etter PD *et al.* (2010) Population Genomics of Parallel Adaptation in
632 Threespine Stickleback using Sequenced RAD Tags. *PLoS Genetics*, **6**, e1000862.
633
634 Huang H, Knowles LL (2014) Unforeseen Consequences of Excluding Missing Data from Next-
635 Generation Sequences: Simulation Study of RAD Sequences. *Systematic Biology, in press.*
636
637 Jeffreys H (1961) *Theory of probability*, 3rd ed. Clarendon Press, Oxford, United Kingdom.
638
639 Jombart T (2008) *adegenet*: a R package for the multivariate analysis of genetic markers.
640 *Bioinformatics*, **24**, 1403–1405.
641
642 Knowles LL (2009) Statistical Phylogeography. *Annual Review of Ecology, Evolution, and*
643 *Systematics*, **40**, 593–612.
644
645 Knowles LL, Alvarado-Serrano DF (2010) Exploring the population genetic consequences of the
646 colonization process with spatio-temporally explicit models: insights from coupled ecological,
647 demographic and genetic models in montane grasshoppers. *Molecular Ecology*, **19**, 3727–3745.
648
649 Körner C (2003) *Alpine plant life: functional plant ecology of high mountain ecosystems.*
650 Springer, Berlin, Germany.
651
652 Kuhner MK (2006) LAMARC 2.0: maximum likelihood and Bayesian estimation of population
653 parameters. *Bioinformatics*, **22**, 768–770.
654
655 Landguth EL, Cushman SA, Murphy MA, Luikart G (2010) Relationships between migration
656 rates and landscape resistance assessed using individual-based simulations. *Molecular Ecology*
657 *Resources*, **10**, 854–862.
658

Lanier HC, Massatti, R, He Q, Olson LE, Knowles LL (2015) Colonization from divergent ancestors: glaciation signatures on contemporary patterns of genomic variation in Collared Pikas (*Ochotona collaris*). *Molecular Ecology*, **24**, 3688–3705.

Leuenberger C, Wegmann D (2010) Bayesian computation and model selection without likelihoods. *Genetics*, **184**, 243–252.

Lohse K, Nicholls JA, Stone GN *et al.* (2011) Inferring the colonization of a mountain range-refugia vs. nunatak survival in high alpine ground beetles. *Molecular Ecology*, **20**, 394–408.

Martinkova N, Barnett R, Cucchi T *et al*. (2013) Divergent evolutionary processes associated with colonization of offshore islands. *Molecular Ecology*, **22**, 5205–5220.

Massatti R, Knowles LL (2014) Microhabitat differences impact phylogeographic concordance of codistributed species: genomic evidence in montane sedges (*Carex* L.) from the Rocky Mountains. *Evolution*, **68**, 2833–2846.

Massatti R, Reznicek AA, Knowles LL (2016) Utilizing RADseq data for phylogenetic analysis of challenging taxonomic groups: A case study in *Carex* sect. *Racemosae*. *American Journal of Botany*, **103**, 337–347.

Mevik B-H, Wehrens R (2007) The PLS package: principal component and partial least squares regression in R. *Journal of Statistical Software*, **18**, 1–24.

Morgan K, O'Loughlin SM, Chen B *et al.* (2011) Comparative phylogeography reveals a shared impact of Pleistocene environmental change in shaping genetic diversity within nine *Anopheles* mosquito species across the Indo-Burma biodiversity hotspot. *Molecular Ecology*, **20**, 4533–4549.

Neuenschwander S, Largiader CR, Ray N *et al.* (2008) Colonization history of the Swiss Rhine basin by the bullhead (*Cottus gobio*): inference under a Bayesian spatially explicit framework. *Molecular Ecology*, **17**, 757–772.

Nielsen R, Beaumont MA (2009) Statistical inferences in phylogeography. *Molecular Ecology*, **18**, 1034–1047.

Oaks JR, Sukumaran J, Esselstyn JA *et al.* (2013) Evidence for climate-driven diversification? A caution for interpreting ABC inferences of simultaneous historical events. *Evolution*, **67**, 991–1010.

O'Meara BC, Jackson ND, Morales-Garcia AE, Carstens BC (2015) Phylogeographic Inference Using Approximate Likelihoods. *bioRxiv*, doi: http://dx.doi.org/10.1101/025353.

Papadopoulou A, Knowles LL (2015a) Genomic tests of the species-pump hypothesis: Recent island connectivity cycles drive population divergence but not speciation in Caribbean crickets across the Virgin Islands. *Evolution*, **69**, 1501–1517.

Papadopoulou A, Knowles LL (2015b) Species-specific responses to island connectivity cycles: refined models for testing phylogenetic concordance across a Mediterranean Pleistocene Aggregate Island Complex. *Molecular Ecology*, **24**, 4252–4268.

Papadopoulou A, Knowles LL (2016) Refined hypotheses based on taxon-specific traits in comparative phylogeography. *Proceedings of the National Academy of Sciences, USA, in review*.

Pelletier TA, Carstens BC (2014) Model choice in phylogeography using a large set of models. *Molecular Ecology*, **23**, 3028–3043.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PLoS ONE*, **7**, e37135.

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.

Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW (1999) Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution,* **16**, 1791–1798.

R Core Team (2012) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ray N, Currat M, Foll M, Excoffier L (2010) SPLATCHE2: a spatially explicit simulation framework for complex demography, genetic admixture and recombination. *Bioinformatics*, **26**, 2993–2994.

Shirk AJ, Cushman SA, Landguth EL (2012) Simulating pattern-process relationships to validate landscape genetic models. *International Journal of Ecology*, **2012**, 539109.

Slatkin M (1993) Isolation by Distance in Equilibrium and Non-Equilibrium Populations. *Evolution*, **47**, 264–279.

Wachter GA, Papadopoulou A, Muster C *et al.* (2016) Glacial refugia, recolonisation patterns, and diversification forces in Alpine-endemic *Megabunus* harvestmen. *Molecular Ecology, in press*.

Wegmann D, Currat M, Excoffier L (2006) Molecular diversity after a range expansion in heterogeneous environments. *Genetics*, **174**, 2009–2020.

Wegmann D, Leuenberger C, Neuenschwander S, Excoffier L (2010) ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics*, **11**, 7.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Westergaard KB, Alsos IG, Popp M *et al*. (2011) Glacial survival may matter after all: nunatak signatures in the rare European populations of two west-arctic species. *Molecular Ecology*, **20**, 376–393.

Data Accessibility

The following data, scripts, and other files will be deposited in Dryad: doi:10.5061/dryad.ng3bv

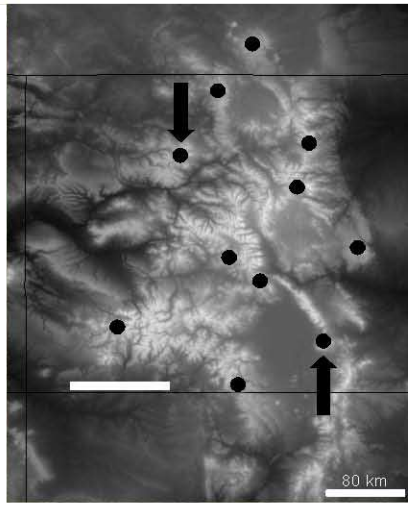
- Species distribution points used in MAXENT
- Empirical datasets (composed of >1000 SNPs for each species)
- Scripts for editing empirical and simulated SNP datasets
- Scripts and settings files used in iDDC analyses

Author Contributions

RM collected specimens, conducted lab work, performed modeling, and wrote the manuscript. LLK helped with the development of the experimental design and the modeling scenarios as well as wrote and edited the manuscript.

767 Table 1. Model statistics for each species and modeling scenario. The Bayes factor represents the ratio between the model with the
768 highest marginal density and the alternative model. Note that the *P*-value refers to the general fit of the data under a model (i.e., it is
769 based on the likelihood of the retained simulated datasets relative to the likelihood of the empirical data, where a high *P*-value
770 indicates the model is capable of generating the data). Parameters include: K_{max}, the carrying capacity of the deme with the highest
771 suitability; m, the migration rate per deme per generation; and N_{Anc}, the ancestral population sizes of initial populations before
772 expansion from refugia. R^2 is the coefficient of determination between a parameter and the six PLSs used herein.
773

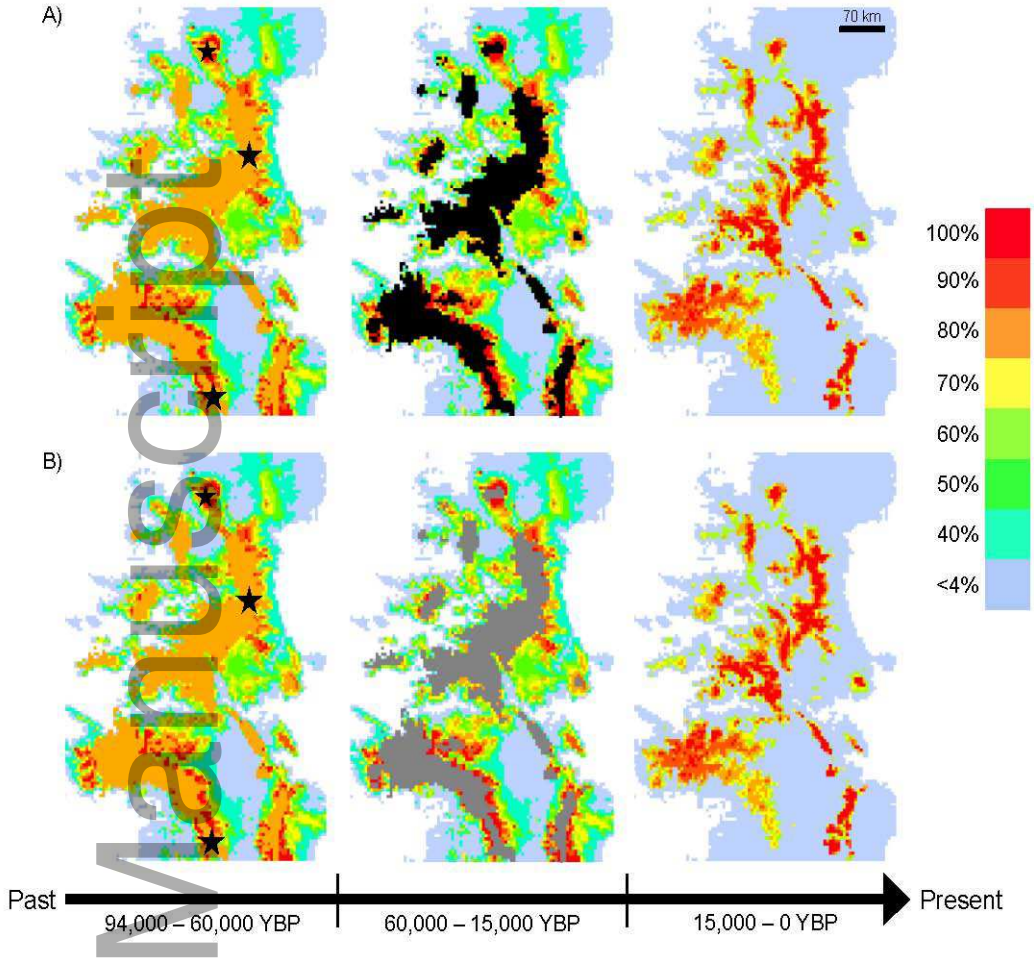
Species	Model	Marginal density	*P*-value	Bayes factor	Parameters	R^2
C. chalciolepis	Barrier	4.87×10^{-5}	0.650	-	K_{max}	0.642
					m	0.966
					N_{Anc}	0.404
	Permeable	1.38×10^{-4}	0.970	2.84	K_{max}	0.698
					m	0.965
					N_{Anc}	0.379
C. nova	Barrier	1.29×10^{-4}	0.844	22.69	K_{max}	0.548
					m	0.961
					N_{Anc}	0.497
	Permeable	5.68×10^{-6}	0.078	-	K_{max}	0.585
					m	0.962
					N_{Anc}	0.479



774

775 Figure 1. Glacial reconstruction for Colorado during the Last Glacial Maximum based on geologic data such as glacial moraines. Note
776 the differential accumulation of glaciers in drainages versus on ridges. Circles represent collecting localities; arrows identify matching

777 localities between the glacial map and the inset (for details, see Table S1). The image is taken from the 'Late Pleistocene glaciers of
778 Colorado' video (Interactive Geology Project, University of Colorado at Boulder, http://igp.colorado.edu/).



779

780 Figure 2. Alternative models used to generate patterns of genetic variation. Both the past and
781 present landscapes were the same between the models, while the intermediate landscape differed
782 in how the glaciated areas were modeled. Specifically, under the barrier model (A), the glaciated
783 areas had a habitat suitability of 0 (denoted by black), whereas under the permeable model (B),
784 the glaciated areas remained permeable (albeit at a reduced carrying capacity, denoted by grey;
785 see text for details). The local carrying capacities of demes differed across the landscapes and
786 were scaled based on habitat suitabilities estimated from ENMs (see colored scale bar). The
787 black stars in the initial landscape mark the locations of ancestral populations used to initiate the
788 simulations.

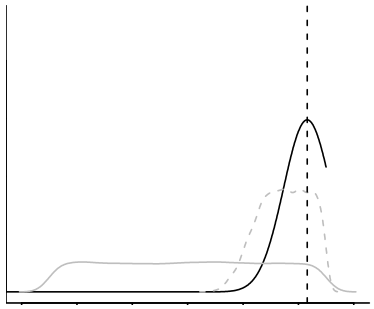
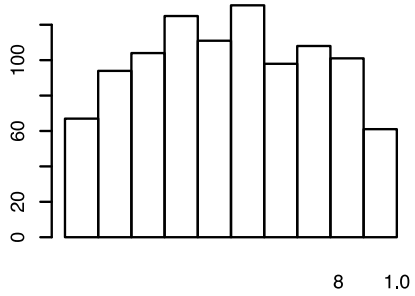


Figure 3. Posterior distribution (black line) and mode (vertical dotted line) of parameter estimates for the most probable model for *C. chalciolepis* (permeable model) and *C. nova* (barrier model). Results are based on a GLM regression adjustment of the 5000 retained simulations. The distribution of the retained simulations (dashed line) and the prior (gray line) demonstrate the improvement that the GLM procedure had on parameter estimates and that the data contained information relevant to estimating the parameters.



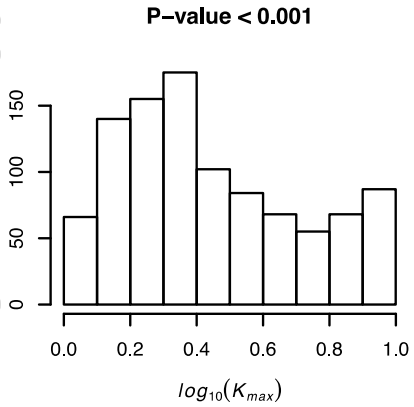


Figure 4. Distribution of posterior quantiles of parameters for *C. chalciolepis* under the permeable model (A) and *C. nova* under the barrier model (B) to evaluate potential bias in the parameter estimates (results are shown only for the most probable models). Bias is measured by a departure from a uniform distribution using a Kolmogorov-Smirnov test (a *P*-value <0.05 indicates a non-uniform distribution). Analyses are based on 1000 pseudo-observations (see text for details). Estimation of N_{Anc} is unbiased while the distributions for *K* and *m* are too wide for both species.